图书情报工作
第62卷 第6期 2018年3月

数字图像语义标注模型比较与分析*

陈金菊 欧石燕
南京大学信息管理学院 南京 210023

摘要:[目的/意义]图像语义标注的基础是图像语义标注模型的构建,对当前主流图像语义标注模型进行梳理和总结,剖析其在图像语义标注中的优缺点,可为后续相关研究提供借鉴和参考。[方法/过程]采用文献调研法,总结出4类主要的图像语义标注模型,即Eakins模型、Jaimes & Chang模型、Kong模型、Panofsky模型。其后采用比较法和归纳法,从语义层次、可扩展性以及应用范围和方式3个方面对前三类模型进行比较分析。[结果/结论]Eakins模型语义层次最全面,语义表达能力最强,应用范围最广;Kong模型的可扩展性最强,适应性最好。

关键词: 图像标注 语义图像标注 图像语义标注模型
分类号: G250
DOI:10.13266/j.issn.0252-3116.2018.06.014

1 引言

近年来,随着数字影像技术和多媒体技术的飞速发展,各种数字图像资源大量涌现,如何对这些海量图像资源进行有效检索和利用,成为一个日益迫切的问题。早期的图像标注主要是通过人工方式选择主题词或关键词对图像内容进行描述,虽然精确度较高,但标注工作量大,标注结果往往具有较强的主观性且稳定性较差。随着计算机技术的发展,基于内容的图像自动标注逐渐成为主流,此类标注主要是从检索角度出发,利用计算机自动提取图像的底层视觉特征(如颜色、形状、纹理),将其与用户的语义查询相匹配,但是由于图像视觉特征并不能完全反映用户的检索意图,导致了语义鸿沟的存在[1]。为了促进数字图像资源的共享与重用,需要对图像内容进行有效的语义标注,增强人们对图像内容的理解。而图像语义标注在很大程度上要依赖于图像语义标注模型。图像语义标注模型是对图像内容进行概念化抽象而得的概念模型,通常采用层次化结构由低到高对图像的视觉特征(颜色、纹理、形状)、逻辑特征(如所含对象、对象间的相对关系)和语义特征(如场景、情感)分层次进行描述,有助于人们更好地理解并描述图像的内容。鉴于图像语义标注模型的这种层次结构,也常被称为图像层次模型。

目前,许多领域(如计算机、生物医学、图书情报等)的学者对图像语义标注尤其是图像语义标注模型进行了研究,为了对国内外图像语义标注研究现状有一个较全面的了解,我们在Web of Science数据库、CNKI数据库和Google Scholar中检索到1985年至2017年间有关图像语义标注模型的中英文文献40余篇,并对其进行了梳理和总结。分析发现,虽然这些文献提及的图像语义标注模型多达20余种,但细究其来源和内容,主要自4个基本模型(Eakins模型[2]、Jaimes & Chang模型[3]、Kong模型[26]和Panofsky模型[29])衍生而来,归属于四大家族,如表1所示。除了这四类主流模型外,还有一些应用范围较小的模型,如M. G. Krause的两层图像内容框架[7],B. Burford等的六层模型[8],Y. Badr和R. Chbeir的两层模型[9]等,由于应用范围过窄,不算作是主流的图像语义标注模型。虽然有关图像语义标注模型的研究成果比较丰硕,但是对这些模型进行全面系统梳理与分析的研究还很少。本文从四个基本的图像语义标注模型出发,对主流的图像语义标注模型进行全面梳理与比较分析,以期为图像语义标注模型的构建和应用提供借鉴与参考。

* 本文系国家社会科学基金重点项目"基于关联数据的学术文献内容语义发布及其应用研究"(项目编号:17ATQ001)和国家社会科学基金重大项目"面向大数据的数字图书馆移动视觉搜索机制及应用研究"(项目编号:15ZDB126)研究成果之一。
作者简介:陈金菊(ORCID:0000-0003-4488-1850),硕士研究生;欧石燕(ORCID:0000-0001-8617-6987),教授,博士生导师,通讯作者,E-mail:oushiyan@nju.edu.cn。
收稿日期:2017-09-16 修回日期:2017-12-02 本文起止页码:116-124 本文责任编辑:易飞

表1　图像语义标注模型分类

类别	基本模型	主要衍生模型
Eakins 模型家族	J. P. Eakins 的图像语义层次模型[2]	• D. Hong 等的三层模型[19] • 于永新的四层模型[20] • 蔡昌许的七层模型[4] • 彭杨的七层模型[21] • 王晓光和徐雷的数字图像语义描述层次模型[40]
Jaimes & Chang 模型家族	A. Jaimes 和 S. F. Chang 的图像语义层次模型[3]	• L. Hollink 等的三层模型[24] • E. K. Chung 和 J. W. Yoon 的三层模型[5] • J. W. Yoon 和 E. K. Chung 的改进三层模型[25] • J. S. Hare 的五层模型[6]
Kong 模型家族	H. Kong 等的图像语义标注模型[26]	• 邓涛等的四层模型[27] • 史婷婷等的三层模型[28]
Panofsky 模型家族	E. Panofsky 的图像语义层次模型[29]	• S. Shatford 的二维模型[30] • N. Conduit 和 P. Rafferty 的二维细化模型[33] • P. Rafferty 和 R. Hidderley 的六层模型[34] • F. Fauzi 和 M. Belkhatir 的五层框架[35]
其他模型	无	• M. G. Krause 的两层图像内容框架[7] • B. Burford 等的六层模型[8] • Badr 和 Chbeir 的两层模型[9]

2　Eakins 图像语义层次模型及其衍生模型

2.1　面向检索的 Eakins 图像语义层次模型

1996 年,英国学者 J. P. Eakins 提出了一个简单且实用的层次化图像语义模型(以下简称"Eakins 模型"),从检索需求的角度首先将图像语义内容自下而上划分为 3 个基本层级(底层特征层、对象层和语义概念层),然后又对每个层级进行更细粒度的划分,如图1(a)所示[10-12]:

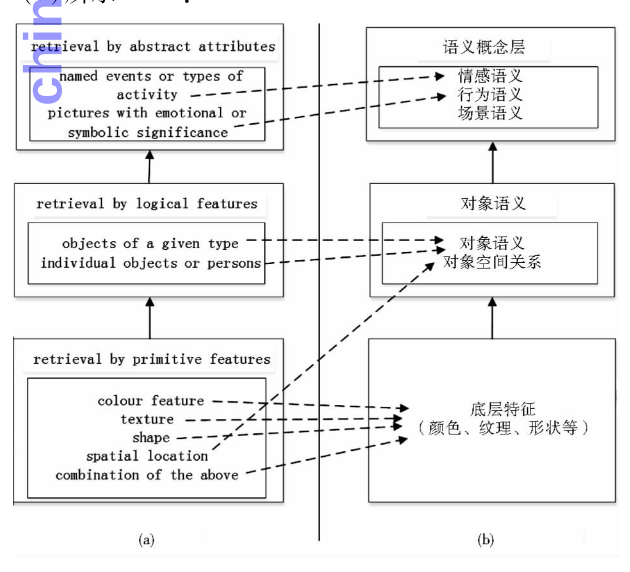


图1　Eakins 模型及对其的重新阐释

原始特征层(Primitive Features):原始特征是指客观的、直接来源于图像本身的纯粹视觉特征,可细分为 5 类,即颜色特征、纹理特征、形状特征、空间位置特征和上述 4 种特征的综合,均不涉及图像的语义内容。面向该层的图像检索不需要参考任何外部知识。基于内容的图像检索多位于该层,被广泛应用于各种专业图像检索,如商标图像注册过程中商标图像的检索[13]。

逻辑层(Logical Features):逻辑特征是指基于图像的视觉特征对图像中的对象进行某种程度的逻辑推理而导出的特征,涉及图像的语义内容。基于逻辑特征的图像检索可进一步细分为:检索特定类型对象的图像、检索单个对象或个人的图像。面向该层的图像检索比面向原始特征层的语义检索更具有普适性。

抽象属性层(Abstract Attributes):抽象属性是指对图像中对象所处场景的目的和意义进行抽象和主观推理所得到的特征。基于抽象属性的图像检索可进一步细分为:检索某一事件或活动的图像、检索具有情感或象征意义的图像。面向该层的图像检索,不仅需要理解图像的语义内容和背景知识,还需具备一定的推理和判断能力。

Eakins 模型是较早的图像语义模型,其提出的目的主要是用于图像检索而非图像的共享与重用。在 Eakins 模型提出之前,美国东卡罗来纳大学和美国路易斯安那大学拉法叶分校的学者 V. N. Gudivada 和 V. V. Raghavan 于 1995 年提出过一个包含原始特征层和逻辑特征层的两层图像语义模型[14],该模型是 Eakins 模型的基础,但由于缺少对象所处场景的抽象与描述,因此语义表达能力明显弱于 Eakins 模型。

2.2 面向语义标注的 Eakins 模型的微调与重释

在 Eakins 模型提出后，国内学者（如武人杰[15]、张婕[16]、陆泉等[17]）对其进行了重新解读、阐释和微调，试图使其能够应用于图像语义标注，相关改进方法基本一致，即将 Eakins 模型重新阐释为如图 1（b） 所示的 3 个层级[15-18]。

底层特征层：该层与 Eakins 模型的原始特征层基本对应，不反映图像的语义内容信息，但去除了原始特征层中的空间位置特征，只包含图像颜色、纹理、形状三类原始特征。

对象层：该层是 Eakins 模型逻辑层和部分原始特征层的综合，细分为对象语义层和对象空间层两个子层次。前者包含图像中涉及的具体对象，与 Eakins 模型的逻辑层相对应；后者指图像中对象间的相互位置关系，与 Eakins 模型原始特征层中的空间位置相对应。

语义概念层：该层与 Eakins 模型的抽象属性层相对应，但突出了对场景语义的描述。该层细分为场景语义层、行为语义层和情感语义层 3 个子层次，其中场景语义层指图像中对象所处的环境；行为语义层指图像中对象的活动行为；情感语义层则是指图像给人的主观感受。

Eakins 模型是一个整体，各层之间存在着依赖关系，中高层语义的获取通常建立在底层特征的基础上，根据先验知识和判断推理得到，但目前要实现高层语义的高效提取还存在一定的困难。重新阐释后的 Eakins 模型与原模型并没有本质上的区别，但是语义层次变得更加清晰、明确，通用性更强，这也是该模型在国内图像语义标注研究中得到广泛应用的主要原因。下文提及的 Eakins 模型均指经过微调和重新阐释的模型。

2.3 Eakins 模型的衍生模型

Eakins 模型创建后，一些研究人员针对特定应用场景，对该模型的语义层次进行增减和调整，产生了许多衍生版本。1998 年，D. Hong 等对 Eakins 模型的子层级进行了局部调整，将图像内容划分为基础视觉内容层、对象内容层和场景内容层三个层次[19]，分别对应于 Eakins 模型的底层特征层、对象语义层和场景语义层。该模型调整的目的是为了针对特定的检索情景灵活地描述图像，但其主要缺陷是将场景视为图像的全局描述，没有考虑行为语义和情感语义，语义表达能力比 Eakins 模型弱。国内学者于永新认为图像语义鸿沟主要体现在对图像中实体间关系描述得不够充分，因此将 Eakins 模型对象层中的两个子层级——对象语义层和对象空间层——提升为与底层特征层和语义概念层相并列的一级层级，衍生出一个四层模型[20]。虽然该模型表达的语义与 Eakins 模型没有本质区别，但是突出了对实体关系的描述。

此外，还有一些研究人员对 Eakins 模型的语义层次进行了扩展，提出了语义更加丰富的多层模型。蔡昌许和彭杨于 2005 年和 2007 年分别提出了各自的七层语义模型，这两个模型的层级基本一致，前六层与 Eakins 模型的 6 个子层相同，但在此基础上增加了第七层，用来表达图像真实的、抽象的更高层的语义，即人们对图像内容的真正理解，通常指图像反映出的真实情景（如婚礼、鸿门宴）[4,21]。与前六层相比，更高层语义层侧重于从图像全局出发，对图像整体的内涵进行描述，抽象度更高。蔡昌许和彭杨的这两个模型虽然没有本质上的区别，但是适用的图像类型有所不同，前者针对一般静态图像，后者针对动画素材图像。

3 Jaimes & Chang 图像语义层次模型及其衍生模型

3.1 Jaimes & Chang 图像语义层次模型

1998 年，A. Jaimes 和 S. F. Chang 提出了一个用于图像自动分类的视觉信息分类框架，该框架自下而上包含区域、感知、对象部件、对象、场景五个层次[22]。2000 年，两人综合运用多领域（如艺术和认知心理学等）的知识将该框架改造为面向图像索引的概念框架（以下简称“Jaimes & Chang 模型”）[3]。在该框架中，图像内容被划分为非视觉和视觉两类，如图 2 所示。非视觉内容是指与图像密切相关但并不直接作为图像一部分的信息，主要包括物理属性、目录信息和相关信息。视觉内容是指观察图像时直接感觉到的信息，自上自下分为呈金字塔结构的 10 个等级：类型技术（type technology）、全局分布（global distribution）、局部结构（local structure）、全局组合（global composition）、一般对象（generic objects）、一般场景（generic scene）、具体对象（specific objects）、具体场景（specific scene）、抽象对象（abstract objects）和抽象场景（abstract scene）[3,23]。其中，前四层属于对图像句法或知觉层面的描述，涉及人或机器所感知的颜色、纹理、元素空间布局等特征，基于内容的图像标注和检索主要关注这四个层次；后六层是对语义或视觉概念的描述，基于语义的图像标注和检索则主要关注这六个层次。与 1998 年提出的视觉信息分类框架相比，该模型更加关注对象语义和场景语义，从一般、具体和抽象三个层面

陈金菊，欧石燕. 数字图像语义标注模型比较与分析[J]. 图书情报工作,2018,62(6):116-124.

对图像中的对象和图像所反映的场景展开更细粒度的描述,但对于不以对象和场景为标注重点的图像适用性并不强。

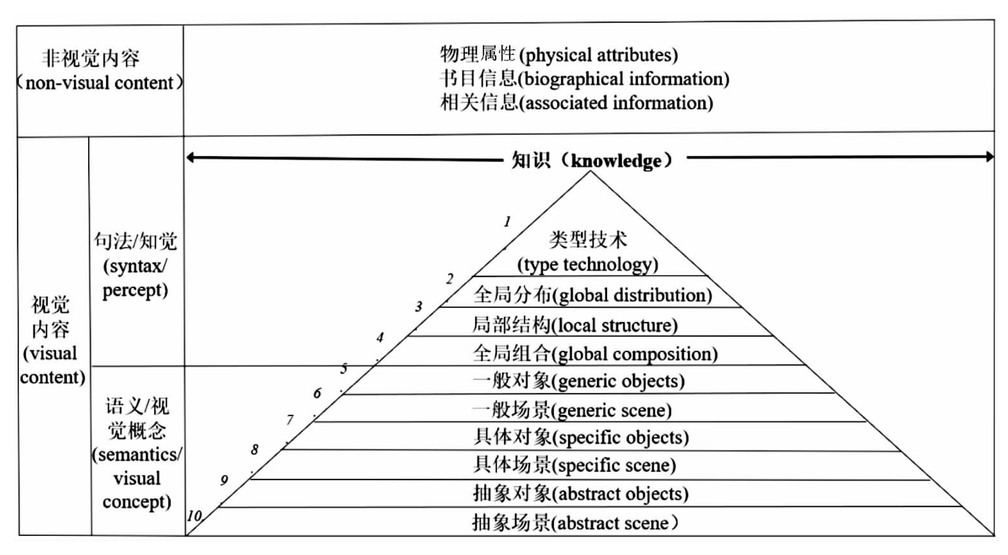


图 2　Jaimes & Chang 图像语义层次模型

3.2　对 Jaimes & Chang 模型的改进和调整

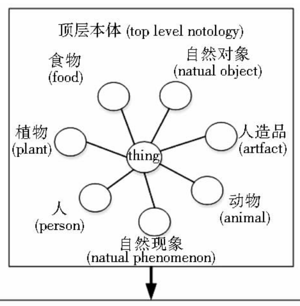
Jaimes & Chang 模型是一个比较通用的语义标注模型,一些学者对该模型进行改进,生成了适用于不同特定应用的衍生模型。2004 年,L. Hollink 等为了解决用户需求与当前图像检索技术不匹配的问题,对 Jaimes & Chang 模型的子层次做了局部增减和调整,使模型的描述粒度变粗,能够囊括更多的语义内容,生成一个自下而上包含概念、知觉和非视觉三个层级的用户图像描述分类框架,其中概念层对应 Jaimes & Chang 模型的语义/视觉概念层并增加了时间、地点和事件,知觉层与 Jaimes & Chang 模型的句法/知觉层大致相对应,但没有进一步划分子层次,非视觉层与 Jaimes & Chang 模型的非视觉内容层大致相对应,主要关注图像的描述性信息,即图像的元数据,如创建者、日期和题名等[24]。该框架没有对 Jaimes & Chang 模型的语义层次进行扩展,只是对该模型的子层次做了局部增减和调整,使模型的描述粒度变粗,能够囊括更多的语义内容。

2011 年,E. K. Chung 和 J. W. Yoon 在深入分析图像检索需求底层结构的基础上,提出了一个图像特征描述框架,自上而下包含 3 个层次:非视觉特征层、语法特征层和语义层[5]。非视觉特征与与 Jaimes & Chang 模型的非视觉内容层大致相对应,但将相关信息扩大到上下文信息,即将相关信息作为上下文信息的子类之一;语法特征层与 Jaimes & Chang 模型的句法/知觉层大致相对应,但删除了类型技术;语义层与 Jaimes & Chang 模型的语义/视觉概念层大致相对应,但增加了人、时间、地点和活动。该模型同样仅是对 Jaimes & Chang 模型内部语义层次的粒度大小和包含的语义内容做增减和微调,但该模型扩展了 Jaimes & Chang 模型的语义/视觉概念层的语义内容,总体的语义表达能力比 Jaimes & Chang 模型有所提升。同年,两人通过研究发现,用自然语言表达的提问能够更好地反映用户的图像检索需求,因此将上述的三层图像特征描述框架又调整为图像需求特征、图像特征和相关信息 3 个层次[25]。其中,图像需求特征是指用户图像检索需求的上下文环境,如检索动机等,图像特征层包含 Jaimes & Chang 模型中的非视觉特征、句法/知觉对象特征和语义/视觉概念特征,相关信息层是将 Jaimes & Chang 模型中非视觉特征的相关信息单独做为一个层级。该模型在 Jaimes & Chang 模型的语义层次的基础上,增加了图像需求特征层,使模型的整体语义表达能力增强,但是也增加了模型的复杂度和标注难度。2006 年,J. S. Hare 等为了弥合图像检索过程中存在的语义鸿沟问题,将语义鸿沟的特征描述为从原始媒体(图像)到媒体内容的全语义(对象关系及其他)理解,提出了五层语义渐变模型,包括原始图像、视觉描述符、对象、对象名称、语义 5 个层级[6],试图缩小语义鸿沟。与 Jaimes & Chang 模型相比,该模型的语义层次划分粒度较粗,没有侧重于对对象和场景的细粒度描述,适用于语义粒度较粗糙的图像标注。

4 Kong 图像语义标注模型及其相关模型

4.1 Kong 图像语义标注模型

2006 年,H. Kong 等从图像中包含的对象着手,对图像内容进行分类,提出了一个可扩展的图像语义标注本体模型(以下简称"Kong 模型")[26]。该模型首先包含一个顶层本体,定义了描述图像中对象类型的 7 个类, 即人、动物、植物、人造品、食物、自然对象、自然现象,提供通用的图像标注框架,较好地囊括了图像对象层的语义内容,但是语义粒度较粗。为了表示对象与背景间以及对象间的空间关系,该模型中还定义了一个空间本体,包含 8 个方向关系和 8 个拓扑关系。此外,为了使图像检索获得更高的查准率,H. Kong 等使用个性化本体对图像进行语义标注,允许用户根据需要在顶层本体基础上建立个性化本体。图 3 所示为顶层本体向个性化本体转化的过程示例,展示如何通过在顶层本体中添加关于特定对象的外部知识实现顶层本体向个性化本体的转化。譬如,要对篮球运动员易建联的图像进行语义标注,用户拥有关于该篮球运动员的许多外部知识,其中之一是"易建联是一位中国男性,他是广东宏远篮球俱乐部的运动员"。用户首先从顶层本体中找到与图中对象"易建联"的国籍和所在俱乐部相对应的类"自然对象"和"人造品",然后通过这些类的具体实例"中国"和"广东宏远"与对象"易建联"建立关联,生成一个针对"易建联"的个性化本体。



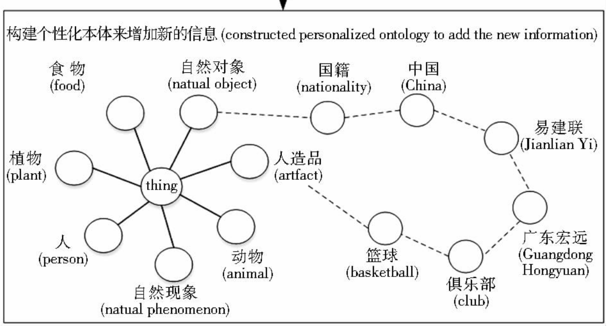


图 3 顶层本体向个性化本体的转化

4.2 Kong 模型的相关模型

一些学者在研究中借鉴了 H. Kong 等以对象为中心的图像内容分类方法和基于个性化本体对对象进行语义标注的思路,构建了一些基于本体的图像语义标注与检索模型。2008 年,邓涛等提出了一个基于本体的图像语义标注与检索模型 ImageQ,其中包含了一个四层的图像内容描述模型。该模型自上而下包含四个层次: 图像元数据,即反映图像外部特征的元数据; 客体对象及其所处背景和所处场景信息; 主体对象及其相关属性信息;主客体对象间的语义关系[27]。与 Kong 模型相比,该模型仍重点关注对象层语义,但是扩展了与对象相关的语义以及图像的外部特征,语义表达能力更强。在该模型中,邓涛等借鉴 Kong 模型的研究思路,针对具体应用领域,仅提供顶层通用本体模型,但允许用户对顶层本体进行更新,包括添加、修改和删除本体中的概念、属性和关系,最终实现领域本体的个性化。2010 年,史婷婷等借鉴邓涛等的上述四层图像内容描述模型,提出了一个三层图像内容描述模型[28]。与邓涛等的四层模型相比,该模型去除了图像元数据特征层,不对图像的外部特征进行描述,只重点关注图像的内部语义特征,但总体上采用了与 H. Kong 等类似的思路,即以对象为中心构建了个性化本体,这种以对象为中心的建模思路很好地增强了图像检索系统的适应性。

5 Panofsky 图像语义层次模型及其衍生模型

5.1 Panofsky 图像语义层次模型

1955 年,E. Panofsky 在对文艺复兴时期的艺术图像进行研究时,提出了一个如图 4 所示的分析模型(以下简称"Panofsky 模型"),包含 3 个层次:前图像志描述(pre-iconography description),指对图像所表达的主题的描述,包括事实和情感;图像志分析(iconography analysis),指对图像中可以识别名称的客观事物的分析;图像学阐释(iconology interpretation),指对图像内涵的阐释[29]。该模型主要关注图像的高层语义信息,不考虑图像的原始物理特征。值得注意的是,该模型只是一个理论分析框架,而非一个具体的语义标注模型,无法直接应用于图像的语义标注。

5.2 对 Panofsky 模型的扩展

为了将 Panofsky 模型应用于具体的图像标注,一些学者对 Panofsky 模型进行了扩展,提出了语义更加

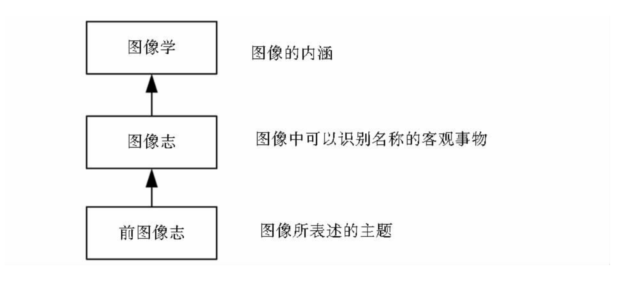


图4　Panofsky 图像分析模型

丰富的模型。1986 年，S. Shatford 通过对 Panofsky 模型的三个语义层次进行横向扩展，提出了一个被称为 Panofsky-Shatford 多面矩阵的二维模型，试图应用于所有类型图像的语义标注[30－31]。该模型包含通用、具体和抽象 3 个层级，分别与 Panofsky 模型的前图像志、图像志和图像学 3 个层次相对应，每个层级下又设 who、what、when 和 where 四个子层级，分别对应图像中的对象、事件(活动)、时间和地点，形成 12 类图像特征，极大地丰富和细化了 Panofsky 模型的语义内容[32]。例如，抽象的地点表示象征性的地点(如天堂)。该模型后来被 N. Conduit 和 P. Rafferty 应用到图像标引，根据图像库中的用户提问和 33 个图像管理员在工作中关注的图像特征，对矩阵做了进一步细化，如将通用类地点细化为室内和室外[33]。这些细化工作使得 Panofsky-Shatford 多面矩阵更加丰富和完善。2007 年，P. Rafferty 和 R. Hidderley 借鉴 Panofsky 等的方法来阐释图像，从图像内容标引的角度，提出了一个六层模型，包含书目信息、结构性内容、整体内容、对象内容、对图像整体的解释以及对其中对象的解释[34]。该模型扩展了图像的外部物理特征，重点关注图像的中高层语义特征，侧重于对对象和情感的描述。2013 年，F. Fauzi 和 M. Belkhatir 借鉴 Panofsky 模型以及其他相关模型，提出了一个以用户为中心的、基于概念的面向自动多面化索引的框架。该框架分析了 Web 图像上下文信息的语义，并将其分为 5 个广泛的语义概念：①信号：指图像的底层视觉特征；②对象：指图像中的实体，分为有生命和无生命两种；③关系：指图像中对象间的关系，以及创作者、图像类型等外部特征关系；④场景：指基于图像包含的所有对象，将图像描述为一个整体；⑤抽象：指图像中表达的抽象概念[35]。该框架将仅包含中高层语义的 Panofsky 模型做了向下语义扩展，增加了对图像底层视觉特征的描述(信号)，并对 Panofsky 模型的图像志做了纵向扩展，增加了对象和关系两个层次。因而，该框架整体语义表达能力更强。

6　模型比较分析

在前文阐述 4 类图像语义标注模型的基础上，本节对上述 Eakins 模型、Jaimes & Chang 模型和 Kong 模型进行深入比较，由于 Panofsky 模型只是一个理论分析框架而非一个具体的模型，因此在此不将其纳入对比分析范围。鉴于每个家族中的衍生模型均数量众多，我们只选取每类模型中的原始基本模型最为分析比较对象，从语义层次、可扩展性及应用范围和方式 3 个方面展开分析。评价图像语义标注模型有两个重要指标：一是语义表达能力，即模型能否将图像所包含的语义完整地表达出来以及其表达的程度；二是适应性，即模型是否能够满足不同用户的需求，提高适应性的一个重要方法是允许用户在应用过程中对模型进行扩展。

6.1　语义层次

图像语义标注模型的语义层次越全面，所能表达的语义就越丰富。图像的内容特征可分为三大类，即物理特征、对象特征和语义特征。其中，物理特征不涉及图像语义内容，而对象特征和语义特征包含 5 种图像语义内容：对象语义、对象空间、场景语义、行为语义和情感语义。Eakins 模型和 Jaimes & Chang 模型均包含了从图像底层物理特征、对象特征到高层语义特征三个层级的图像内容特征，由低到高表达越来越抽象的语义。但两者包含的语义特征数量不同，Eakins 模型包含 5 种主要语义特征(对象语义、对象空间、场景语义、行为语义和情感语义)，语义层次最全面，所表达的语义也最充分；Jaimes & Chang 模型只包含 3 种主要的语义特征(对象语义、场景语义和情感语义)，缺少对对象空间和行为语义的描述，所表达的语义的全面性和充分性弱于 Eakins 模型。Kong 模型不考虑底层物理特征和高层语义特征，直接从中间层对象特征着手，直接对图像中描述的对象及对象的空间关系进行描述。该模型包含两种对象特征，即对象语义和对象空间，其语义层次的全面性和语义表达的充分性均弱于 Eakins 模型和 Jamies & Chang 模型。如表 2 所示：

表 2　图像语义标注模型的语义层次

语义层次	Eakins 模型	Jaimes & Chang 模型	Kong 模型
对象语义	P	P	P
对象空间	P		P
场景语义	P	P	
行为语义	P		
情感语义	P	P	

注："P"表示该模型包含相应的语义层次

综上所述,Eakins 模型的语义最完整且表达能力最强。人类根据自己对图像的理解和认知获取的图像语义,往往是图像的对象语义和高层语义,上述核心模型均将这一因素考虑在内,但是都缺少对图像所渲染的氛围等其他更抽象语义的描述,在语义概念层的基础上,还有可进一步扩展的空间。

6.2 可扩展性

图像语义标注模型的可扩展性对图像检索系统的生命力和适应性有着重要的影响。我们对 3 个核心模型的可扩展性进行总结,将其划分为强、较强和弱 3 个等级,如表 3 所示。这 3 个核心模型均允许用户对模型进行扩展,其中,Eakins 模型支持对语义层次的增减和语义层级的微调,Kong 模型支持在对象语义层增加子层级,这使得两者均具有良好的适应性, Jimes & Chang 模型虽然也允许用户对模型进行扩展,主要支持语义层次和语义层级的增加,但不支持在应用过程加入外部知识,因此可扩展性和适应性相对较弱。

此外,Eakins 模型和 Kong 模型均允许在图像语义标注过程中加入外部知识(如相关的上下文信息)对图像进行标注,以丰富图像的语义,而不是仅局限于图像中反映的内容。Kong 模型还采用了本体技术,允许用户在顶层本体的基础上,根据领域知识构建个性化本体,进一步增强了模型的可扩展性和适应性。

将本体技术运用于图像语义标注主要有以下几点优势:本体是一种标准化、规范化的知识表示方式,运用本体进行领域建模,提供了领域内统一的概念和关系,增强了检索的准确率[36];本体对概念以及概念间的关系进行了详细的描述,在分散孤立的图像间建立联系,增强了图像间的耦合度[27];本体提供了独立于特殊对象的语义描述手段,有助于实现语义信息的共享和重用[37-38];本体的推理功能有助于实现图像的智能检索。针对不同的应用,可以使用不同的本体来描述图像,例如图像底层特征可以使用 VDO(visual descriptor ontology)本体来描述,该本体包含 MPEG-7 的视觉描述符和对象视觉特征的概念和属性[39];而中高层语义信息通常需要定义特定领域的领域本体来描述。鉴于 Kong 模型的可扩展性以及可基于外部知识进行个性化定制的特性,该模型的可扩展性和适应性最好。

表 3 图像语义标注模型的可扩展性

模型名称	应用中是否可参考外部知识	可扩展性
Eakins 模型	是	较强
Jaimes & Chang 模型	否	较弱
Kong 模型	是	强

6.3 应用范围与应用方式

3 个核心模型都可用于图像语义标注和图像语义检索,但都没有对适用的图像类型和应用领域做出具体的限制,我们通过对模型描述的语义信息进行分析,总结了 3 个核心模型适用的图像特点和应用情景,如表 4 所示。3 个图像语义标注模型,但从模型描述的语义信息可以看出,这 3 个核心模型的应用范围还是有所区别。

表 4 图像语义标注模型的应用范围

模型名称	标注特点	应用领域
Eakins 模型	语义内容全面、多样,主要包括对象、对象空间、行为、场景、情感语义	如艺术(美术、书法等)、历史等领域的图像语义标注
Jamies & Chang 模型	以对象和场景语义为主,一般不包含情感语义	如建筑设计、地理等领域的图像标注
Kong 模型	以对象和对象空间语义为主,一般不包含情感语义	如生物医学、医疗健康等领域的图像标注

Eakins 模型主要用于图像检索,旨在改善图像检索系统的性能,提高检索系统的查准率。该模型能较完整地表达图像中包含的对象、对象空间、行为、场景和情感语义,适用但不限于艺术(美术、书法等)、历史等领域的图像语义标注。王晓光和徐雷等于 2014 年提出了一个敦煌壁画数字图像语义描述层次模型[40],该模型基于 Eakins 模型,通过加入适用于敦煌壁画描述的术语表和图像元数据揭示图像的高层语义,实现了对敦煌壁画数字图像的语义标注。2017 年,徐雷和王晓光以他们提出的数字图像语义描述层次模型为基础,结合叙事型图像包含的情节语义信息,对叙事型图像中的情节语义进行数据建模,实现了该类图像的语义标注与检索[41]。叙事型图像包含的语义内容信息较全面,基本上可以囊括对象、对象空间、行为、场景和情感语义。

Jaimes & Chang 模型主要用于以检索为目的的索引和图像描述的分类。2001 年, C. Jörgensen 等对 Jaimes & Chang 金字塔模型进行探索性评价,通过将该模型用于图像语义描述和语义标引来对其进行验证,结果表明该模型的功能十分强大,不仅可以描述用于检索的视觉内容,指导索引过程,还可以对手工或自动获取的描述进行分类,较好地涵盖了用户描述和标注

图像过程中涉及的图像特征[42]。Jaimes & Chang 模型侧重于对图像的对象语义和场景语义的细粒度描述,因而对以对象和场景为主的图像具有较好的描述能力适用于建筑设计、地理等领域的图像标注。但因为该模型不包含情感语义,不适于蕴含情感的图像的描述。

Kong 模型也主要是用于图像检索,一般从图像对象层语义着手来构建模型,借助本体技术构建顶层本体,并允许用户根据需求构建个性化本体。邓涛和史婷婷等通过在图像检索系统中进行实验,结果表明,基于个性化本体的图像语义标注和检索与其他检索方式(如借助搜索引擎的百度图片搜索)相比,查准率得到了提高[33－34]。Kong 模型是一个对象层语义标注模型,不考虑图像物理特征和语义特征。该模型对以客观对象为主要内容的图像具有较强的描述能力,但一般不包含情感语义,适用但不限于生物医学、医疗健康等领域的图像标注。

7 结论

图像语义标注模型是图像语义标注和语义检索的前提与基础,为其提供了一种描述图像内容(含底层视觉特征和语义特征)的框架。本文对 4 类主要的图像语义标注模型(Eakins 模型、Jaimes & Chang 模型、Kong 模型、Panofsky 模型)及其衍生模型进行了分析和总结。在这 4 个模型中,除 Panofsky 模型是一个抽象图像语义分析框架,其余三个模型都是可以具体应用的模型。因此,我们从语义层次、可扩展性以及应用范围和方式 3 个方面对这三类模型进行比较分析。从语义表达能力上来说，Eakins 模型的语义层次最全面,语义表达能力最强；从适应性上来说，Kong 模型的适应性最好,不仅标注过程参考了外部知识,而且允许用户在应用过程中对模型进行扩展,根据自己的专业知识和需求构建个性化本体。从应用范围上来说,Eakins 模型应用最广泛,许多研究人员以该模型为基础,相继提出了许多改进模型和基于此模型的相关应用,使该模型得到了广泛认可。其他两类模型虽然也得到较广泛的应用,但是相对于 Eakins 模型影响力较弱。

参考文献:

[1] SMEULDERS A W M, WORRING M, SANTINI S, et al. Content-based image retrieval at the end of the early years[J]. IEEE transactions on pattern analysis and machine intelligence, 2000, 22(12):1349－1379.
[2] EAKINS J P. Retrieval of still images by content[M]. Lectures on information retrieval. Springer, Berlin, Heidelberg, 2000: 111－138.
[3] JAIMES A, CHANG S F. A conceptual framework for indexing visual information at multiple levels[J]. Proceeding of SPIE-The International Society for Optical Engineering. San Jose: IS&T/SPIE Internet imaging, 2000, 3964: 2－16.
[4] 蔡昌许. 基于语义的图像标注与检索系统研究[D]. 武汉:武汉大学,2005.
[5] CHUNG E K, YOON J W. Image needs in the context of image use: an exploratory study[J]. Journal of information science, 2011, 37(2): 163－177.
[6] HARE J S, LEWIS P H, ENSER P G B, et al. Mind the gap: another look at the problem of the semantic gap in image retrieval [J]. Multimedia Content Analysis Management & Retrieval, 2006, spie v.
[7] KRAUSE M G. Intellectual problems of indexing picture collections [J]. Audiovisual librarian, 1988, 14(2): 73－81.
[8] BURFORD B, BRIGGS P, EAKINS J P. A taxonomy of the image: on the classification of content for image retrieval[J]. Visual communication, 2003, 2(2): 123－161.
[9] BADR Y, CHBEIR R. Automatic image description based on textual data[M]//Journal on data semantics VII. Berlin, Heidelberg: Springer, 2006.
[10] EAKINS J P. Automatic image content retrieval-are we getting anywhere? [C]//Proceeding of Third International Conference on Electronic Library and Visual Information Research. De Mont fort University. Milton Keynes: Aslib,1996: 123－125.
[11] EAKINS J P. Design criteria for a shape retrieval system[J]. Computers in industry, 1993, 21(2): 167－184.
[12] EAKINS J P, GRAHAM M E, BOARDMAN J M, et al. Retrieval of trade mark images by shape feature[C]//Proceeding of first International conference on electronic library and visual information system research. Milton Keynes:De Montfort University, 1996:101－109.
[13] PETKOVIC D. Query by image content[C]//Oral presentation to storage and retrieval for image and video databases. California: San Jose,1996.
[14] GUDIVADA V N, RAHAVAN V V. Content-based image retrieval systems[J]. IEEE computer, 1995, 28(9): 18－22.
[15] 武人杰. 图像层次语义描述的初步研究[J]. 电脑开发与应用,2011(5):12－14.
[16] 张捷. 图像语义标注[J]. 电脑开发与应用,2012(1):10－12.
[17] 陆泉,丁恒. 基于情感的图像检索研究综述[J]. 情报理论与实践,2013(2):119－124.
[18] 黄质纯. 基于语义的图像检索及相关技术的研究[D]. 广州:华南理工大学,2012.
[19] HONG D, WU J, SINGH S S. Refining image retrieval based on context-driven methods[C]//Storage and retrieval for image and video databases VII. 1998: 581－592.
[20] 于永新. 基于本体的图像语义识别和检索研究[D]. 天津:天津大学,2009.
[21] 彭杨. 基于本体的动画素材图像语义标注研究[D]. 长沙:湖

南师范大学,2009.
[22] JAIMES A, CHANG S F. Model-based classification of visual information for content-based retrieval[C]// Proceedings of SPIE - The International Society for Optical Engineering. 1998:402 - 414.
[23] TOUSCH A M, HERBIN S, AUDIBERT J Y. Semantic hierarchies for image annotation: a survey[J]. Pattern recognition, 2012, 45(1): 333 - 345.
[24] HOLLINK L, SCHREIBER A T, WIELINGA B J, et al. Classification of user image descriptions[J]. International journal of human-computer studies, 2004, 61(5): 601 - 626.
[25] YOON J W, CHUNG E K. Understanding image needs in daily life by analyzing questions in a social Q&A site[J]. Journal of the Association for Information Science & Technology, 2011, 62(11): 2201 - 2213.
[26] KONG H, HWANG M, KIM P. The study on the semantic image retrieval based on the personalized ontology[J]. International journal of information technology, 2006, 12(2): 35 - 46.
[27] 邓涛,郭雷,杨卫莉. 基于本体的图像语义标注与检索模型[J]. 计算机工程,2008(17):188 - 190.
[28] 史婷婷,闫大顺,沈玉利. 基于个性化本体的图像语义标注和检索[J]. 计算机应用,2010(1):90 - 93.
[29] PANOFSKY E. Meaning in the visual art: papers in and on art history[M]. New York:Doubleday Anchor Books, 1955:39 - 40.
[30] SHATFORD S. Analyzing the subject of a picture: a theoretical approach[J]. Cataloging & classification quarterly, 1986, 6(3): 39 - 62.
[31] 黄崑,王珊珊,耿骞. 国外图像特征研究进展与启示[J]. 图书情报工作,2015,59(8):138 - 146.
[32] CHOI Y, RASMUSSEN E M. Searching for images: the analysis of users' queries for image retrieval in American history[J]. Journal of the Association for Information Science and Technology, 2003, 54(6): 498 - 511.
[33] CONDUIT N, RAFFERTY P. Constructing an image indexing template for the children's society: users' queries and archivists' practice[J]. Journal of documentation, 2007, 63(6): 898 - 919.
[34] RAFFERTY P, HIDDERLEY R. Flickr and democratic indexing: dialogic approaches to indexing[J]. Aslib Proceedings, 2007, 59(4/5): 397 - 410.
[35] FAUZI F, BELKHATIR M. Multifaceted conceptual image indexing on the World Wide Web[J]. Information processing & management, 2013, 49(2): 420 - 440.
[36] 张梅,郝佳,阎艳,等. 基于本体的知识建模技术[J]. 北京理工大学学报,2010(12):1405 - 1408,1431.
[37] 张杨,房斌,徐传运. 基于本体和描述逻辑的图像语义识别[C]//南宁:全国安全关键技术与应用学术会议. 2009.
[38] BRACHMAN R J, SCHMOLZE J G. An overview of the KL-ONE knowledge representation system[J]//Cognitive science, 1985, 9(2): 171 - 216.
[39] SIMOU N, TZOUVARAS V, AVRITHIS Y, et al. A visual descriptor ontology for multimedia reasoning[C]//Proceedings of workshop on image analysis for multimedia interactive services. Montreux, 2005: 13 - 15.
[40] 王晓光,徐雷,李纲. 敦煌壁画数字图像语义描述方法研究[J]. 中国图书馆学报,2014,40(1):50 - 59.
[41] 徐雷,王晓光. 叙事型图像语义标注模型研究[J]. 中国图书馆学报,2017,43(5):70 - 83.
[42] JORGENSEN C, JAIMES A, BENITEZ A B, et al. A conceptual framework and empirical research for classifying visual descriptors[J]. Journal of the Association for Information Science and Technology, 2001, 52(11): 938 - 947.

作者贡献说明:

陈金菊:撰写并修改论文;

欧石燕:提出研究方向并拟定研究要点、修改论文。

Comparison and Analysis of the Semantic Models for Digital Image Annotation

Chen Jinju Ou Shiyan
School of Information Management, Nanjing University, Nanjing 210023

Abstract: [**Purpose/significance**] Semantic annotation of digital images is an effective way to solve this problem. The foundation of semantic image annotation is the construction of semantic models. This paper intends to review the existing mainstream semantic models for image annotation, and explore their advantages and disadvantages. [**Method/process**] Firstly, four representative semantic models for image annotation were reviewed, including Eakins model, Jaimes & Chang model, Kong model and Panofsky model, using literature survey, and then the first three models from three aspects (i. e. semantic level, extensibility and application range) were compared and analyzed using comparative analysis. [**Result/conclusion**] Through the above analysis, it can be concluded that Eakins model has the most comprehensive semantic level, the strongest semantic expression ability and the widest application range, whereas Kong model is the most scalable and adaptable one.

Keywords: image annotation semantic image annotation semantic models for image annotation